Sensitivity–Specificity Tradeoff
(ATTRIB detection)

GPT-5.2 (fewshot)

DeepSeek R1 (fewshot)

Claude Opus 4.5 (zeroshot)

Sensitivity (Recall on post-1989)

Specificity (1 − FP rate on pre-1989)